Velocity estimation by simulated annealing: problems and prospects

Daniel Rothman

INTRODUCTION

This paper proposes a technique for estimating laterally-variable interval velocities in a 2-D Earth without making a first guess. The proposed method would obtain only approximate results; greater accuracy can be achieved by using the result of this technique as the first guess in a linearized approach to velocity inversion.

I propose a Monte Carlo approach similar to the algorithm I used to estimate statics corrections in Rothman (1985). Although the computational workload for velocity inversion appears at first to be unmanageably greater than it is for the statics problem, a few key simplifications might make the solution to the velocity problem computable.

This proposal investigates two applications. For simplicity, I first examine the problem of borehole tomography. I look at it first as a problem of traveltime inversion, but I speculate briefly on the use of wave-theoretic methods. Next, I look at the reflection seismic problem, treating velocity inversion as a problem of stack optimization (Toldi, 1985). In all cases the issue of repeated forward modeling is an important problem. I identify the key issues, and propose possible solutions.

BOREHOLE TOMOGRAPHY

Inversion of traveltimes: the ray-tracing problem

I begin with the inversion of traveltimes because the required computational effort is strongly diminished by the convenience of working with traveltimes, rather than the seismograms themselves.

Consider a rectangular portion of a 2-D Earth, with boreholes drilled on opposite sides of the rectangle. On the left side are a line of sources, and on the right are a line of receivers. In separate experiments, each source excites seismic waves which propagate through the 2-D Earth to the opposite side, where the transmitted waves are recorded by the receivers in the other borehole. The time of each direct arrival is then observed on each seismogram; the dataset to be inverted is then this set of observed traveltimes t_{ij}^{obs}. This problem, known as *borehole tomography,* is discussed by Al-Yahya (1985) and Worthington (1984).

Divide the rectangular Earth into M rows and N columns of blocks. The slowness (the reciprocal of seismic velocity) of the block in the k th row and the l th column is denoted by w_{kl}. The problem is to find the slowness model $\mathbf{w} = \{w_{kl}\}$ that yields the calculated traveltimes $\{t_{ij}^{calc}\}$ which best fit the observables $\{t_{ij}^{obs}\}$. Expressing this as a least-squares problem, I write

$$\min_{\mathbf{w}} E(\mathbf{w}) \tag{1}$$

where

$$E(\mathbf{w}) = \sum_{ij} [t_{ij}^{calc}(\mathbf{w}) - t_{ij}^{obs}]^2 \tag{2}$$

The usual approach to the solution of the optimization problem (1) is to specify an initial model $\mathbf{w}_0$ that approximates the (as yet unattained) solution. The problem is to now find the perturbation $\Delta\mathbf{w} = \mathbf{w} - \mathbf{w}_0$ that solves the optimization problem

$$\min_{\Delta\mathbf{w}} E'(\Delta\mathbf{w}) \tag{3}$$

where

$$E'(\Delta\mathbf{w}) = \sum_{ij} [\, t_{ij}^{calc}(\mathbf{w}_0+\Delta\mathbf{w}) - t_{ij}^{obs}\,]^2 \ . \tag{4}$$

Perturbations are used because they arise naturally when the problem is linearized; if the perturbations are small enough such that first-order effects adequately characterize the change in the model, then new traveltimes can be recalculated along the old, unperturbed raypaths (Aki and Richards, 1980). If the perturbations are large enough such that first-order perturbation theory does not apply, new rays must be traced to obtain a new set of t_{ij}^{calc} for each new model $\mathbf{w}$.

Assume for the moment that a new set of t_{ij}^{calc} can be easily computed, regardless of the size of the perturbation in slowness. In analogy with the statics algorithm described in Rothman (1985), the algorithm for velocity estimation that I propose here updates one parameter at a time, each time choosing the new value for that parameter from a probability distribution derived from $E(\mathbf{w})$.

Let $\mathbf{w}^{kl}$ be an interval slowness model that is equal to a previously estimated slowness model $\mathbf{w}$ everywhere except possibly at w_{kl}. The new value for w_{kl} is chosen from the probability distribution

$$P(\mathbf{w}^{kl}) = \frac{\exp\{-E(\mathbf{w}^{kl})/T\}}{\sum\limits_{\mathbf{w}^{kl}} \exp\{-E(\mathbf{w}^{kl})/T\}} , \qquad (5)$$

where T is of course the control parameter called "temperature."

The key computational issue here is the calculation of the probability distribution (5). To evaluate each $E(\mathbf{w}^{kl})$, a new set of $t_{ij}^{calc}(\mathbf{w}^{kl})$ must be calculated. In principle, this means that rays must be traced through the entire model, from each source to each receiver, for each possible value of w_{kl}. Specifically: if there are S sources and R receivers, and each w_{kl} may assume one of W values, then there are $S \times R \times W$ rays to trace to form the probability distribution (5). But this is the work needed to update only one parameter. A single iteration of the algorithm requires that $N \times M$ parameters be updated; thus the number of rays to trace for each iteration is a whopping $S \times R \times W \times N \times M$. What a problem!

Before discussing how the rays would be traced, it is worth making the obvious observation that all rays do not pass through all blocks. Indeed, only a small fraction of the rays pass through a given block. Thus rays need be retraced only if they go through the block of interest. One of the pitfalls with this reasoning, however, is that one can never be sure which rays pass through which blocks until the rays are traced. But a scheme could probably be developed that would incorporate, say, 90% of the relevant rays each time. The problem of repeated ray tracing still remains, however.

Inversion of traveltimes: the ray-tracing solution

The ray-tracing problem can probably be solved, but only after a reformulation and simplification of the problem. I identify some key issues below, and then propose a solution.

The objective of velocity estimation by simulated annealing is to obtain solutions that are independent of the first guess. Although first-guess-independence seems to preclude the use of old raypaths to obtain new raypaths and traveltimes through a (non-linearly) perturbed model, this is not necessarily the case.

One method that uses original rays to obtain new rays and traveltimes without invoking first-order perturbation theory is *dynamic ray tracing* (Cerveny et al., 1984). Dynamic ray tracing provides a computationally efficient technique for two point ray tracing. In the problem at hand, only traveltimes are needed, not the rays themselves. Cerveny et al. (1984) show how the traveltimes for a new, nearby raypath can be obtained from the old ray. Their theory is approximate, but not based on any linearizations; it is based instead on the paraxial approximation to the wave equation. Because the new ray must be close (in an asymptotic sense) to the old ray, and because the paraxial approximation can only be iterated a finite number of times, dynamic ray tracing does not appear to eliminate the need to retrace rays. It should, however, be able to considerably lessen the computational load that an inefficient "shooting" algorithm would require. Unfortunately, this computational saving would probably be insufficient. Even if dynamic ray tracing were used, the basic problem of too much retracing of rays would remain.

The solution to this ray-tracing problem demands a compromise. Although first-order perturbation theory appears inappropriate for a nonlinear problem, it can still be used advantageously. Consider limiting the possible perturbations of a single block so the perturbations are *always* small enough such that first-order perturbation theory holds. (The allowable linear perturbations may in fact encompass all possible slowness values of interest. Because only one block changes at a time, the perturbation of the entire model is small.) At the start of each iteration, new rays would be traced through the current velocity model. This would mean tracing (only) $S \times R$ rays. Using these rays, each slowness block w_{kl} would be updated by choosing a random number from the probability distribution (5). New traveltimes would be computed by calculating the traveltimes along the old raypaths in which only the slowness w_{kl} is different. Each slowness block would be updated simultaneously at the end of the iteration, *not sequentially* at the time of calculation. The next iteration begins by tracing new rays once again. This is a mixture of linear and nonlinear methods, and is somewhat reminiscent of the usual iterative linearization of a nonlinear problem. But there is one important difference: the result of each iteration does not necessarily decrease the objective function (1); after a sufficient number of iterations the initial guess becomes insignificant.

This combination of linear and nonlinear techniques is the algorithm that would be written if the Monte Carlo algorithm were run in parallel. Normally, because I do not use a parallel computer, I describe simulated annealing as an algorithm that sequentially updates each parameter. As I've noted elsewhere (Rothman, 1984), simulated annealing can be implemented on a parallel machine; indeed, it *should* be. If the algorithm were executed in parallel, equation (5) would be evaluated for each slowness parameter simultaneously, and each parameter would also be updated simultaneously. Thus rays would (and could) only be retraced at the end of each iteration.

Viewed from this parallel perspective, is the solution now computable? Based on my experience with statics estimation, I would expect that several thousand iterations would be needed. The question then reduces to whether rays can be traced several thousand times. The answer is a qualified yes. If the number of sources and receivers is small, simple ray tracing that yields only traveltimes (nothing else is required) should be inexpensive enough to repeat many times.

Inversion of wavefields

It is attractive to consider using waves instead of rays. Waves are easier than rays to program, easier to understand, and they yield more information. Unfortunately, a wave-theoretical approach is not nearly as conducive to simplification as a ray-theoretical method. Unlike rays, waves are global, rather than local, phenomena. Thus waves appear impractical for use in a nonlinear, first-guess-independent algorithm that updates one parameter at a time. But there are some conceptual advantages worth considering. I will first describe how waves would be used in the Monte Carlo algorithm I outlined above. I will then show how this perspective may shed some light on the problem of *linearized* velocity estimation.

I would try the following technique. Each wavefield would be backpropagated from the receivers back to the relevant source. The best slowness model is then that model which maximally focuses the wavefield at the source. For source s_i, call this focusing measure $F(s_i, \mathbf{w})$. F is a norm that is zero for no focusing and one for perfect focusing. Then set

$$E(\mathbf{w}) = -\sum_i F(s_i, \mathbf{w}) \tag{6}$$

to be the objective function. The best estimate of the slowness model $\mathbf{w}$ is found at the global minimum of $E(\mathbf{w})$.

For each (nonlinear) perturbation of velocity, a new wavefield would have to be computed. Thus, there would be $S \times N \times M \times W$ wavefields to compute for each iteration, an impractically large number. There is a small consolation, however. As the slowness block w_{kl} gets closer to the sources, the backpropagated wavefield is easier to compute—it need be extrapolated only over a short distance. But there are still too many computations to perform. Thus a wave-theoretical approach is probably not practical.

The idea of backpropagation, focusing, and the minimization of the quantity in equation (6) deserves further attention, however. It would probably be an excellent tool for linearized inversion of the velocity field, in the conventional context in which one makes perturbations about a first guess. Starting with an initial slowness model $\mathbf{w}_0$, one would seek, in analogy to equation (4), the best first-order perturbation $\Delta\mathbf{w}$ that minimizes

$$E'(\Delta \mathbf{w}) \;=\; -\sum_i F(s_i, \mathbf{w}_0 + \Delta \mathbf{w}) \tag{7}$$

The ability to minimize (7) depends on the ability to calculate the partial derivatives $\partial E' / \partial \Delta w_{kl}$, which should be possible in a practical context. This linearized wave-theoretical approach is dependent on an initial guess, but it nevertheless appears to be an interesting avenue for study in itself. Devaney and Oristaglio (1984) discuss some related ideas.

VELOCITY ESTIMATION FROM REFLECTION DATA

I would follow Toldi (1985) and seek the set of interval velocities that produce the optimally stacked data; i.e., the stack with the greatest power.

Parameterize a 2-D Earth by slownesses $\{w_{kl}\}$, where k denotes depth and l denotes lateral location. Designate by $d_{yh}(t)$ the trace at midpoint y and offset h. Assume that all reflectors are flat. For each slowness model $\mathbf{w}$, there exists a set of traveltime corrections that transform a non-zero-offset trace to a zero-offset trace. These traveltime corrections produce data $d_{yh}[\tau(\mathbf{w})]$, where $\tau(\mathbf{w})$ is the transformation to zero-offset time. This traveltime correction is *not* hyperbolic NMO correction.

The objective function is now

$$E(\mathbf{w}) \;=\; \sum_y \sum_\tau \left(\sum_h d_{yh}[\tau(\mathbf{w})] \right)^2 . \tag{8}$$

New estimates of each slowness w_{kl} are obtained by choosing a random number from the probability distribution (5), in which $E(\mathbf{w}^{kl})$ is the stack power obtained from a slowness model that is the same as $\mathbf{w}$ everywhere except at w_{kl}. There are two computational issues in this problem: the calculation of $\tau(\mathbf{w})$ and the power computations inside the parentheses in (7).

To compute the traveltime corrections $\tau(\mathbf{w})$, I would trace rays in the same way that I advocated for the borehole tomography problem. Thus I would compute new traveltimes using old raypaths, but I would retrace the rays after each iteration. Unfortunately, a new ray would have to be retraced for each value of τ, making these computations considerably more expensive than the computations required in borehole tomography. The assumption of flat reflectors should facilitate the recalculation of raypaths. I do not know how crucial this assumption is, but it does appear that dip must be specified.

The stack-power computations really shouldn't present a major problem. The traces should be interpolated (in time) before commencing the algorithm, by, say, a factor of four or so. Then the traveltime corrections to $d_{yh}(t)$ could be made by nearest-neighbor interpolation.

CONCLUSIONS

Velocity estimation by simulated annealing requires repeated forward modeling. Because the Monte Carlo algorithm updates one parameter at a time, the local nature of ray-theoretical techniques makes ray tracing more useful than wave-based methods.

If the perturbations of velocity (or slowness) models are small enough such that first-order effects adequately characterize the change in the model, then new traveltimes can be computed using the old, unperturbed raypaths. Although a first-guess-independent method may appear to preclude the use of first-order perturbation theory, this theory can be profitably applied to the estimate of each velocity (or slowness) parameter when the rest of the model is held constant. New rays then need only be retraced once per iteration. This technique is appropriate for parallel computations, but it may of course be implemented sequentially.

ACKNOWLEDGMENTS

I thank Chuck Sword and Rob Haar for many interesting discussions. I also thank Ivan Psencik, not only for pointing out a possible application of dynamic ray tracing, but also for his hospitality during a recent visit I made to Prague.

REFERENCES

Aki, K. and Richards, P., 1980, Quantitative seismology, W.H. Freeman and Co.

Al-Yahya, K., 1985, An iterative solution to seismic tomography, Stanford Exploration Project Rep. 42, 129-150.

Cerveny, V., Klimes, L., and Psencik, I., 1984, Paraxial ray approximations in the computation of seismic wavefields in inhomogeneous media, Geophys. J. R. astr. Soc., **79**, 89-104.

Devaney, A.J. and Oristaglio, M.L., 1984, Geophysical diffraction tomography, Presented at the 54th Meeting of the SEG, Atlanta.

Rothman, D.H., 1985, Automatic estimation of very large residual statics, Stanford Exploration Project Rep. 42, 225-264.

Rothman, D.H., 1984, Nonlinear inversion, simulated annealing, and residual statics estimation, Stanford Exploration Project Rep. 41, 297-326.

Toldi, J., 1985, Velocity analysis without picking, Ph.D. thesis, Stanford Exploration Project Rep. 43.

Worthington, M.H., 1984, An introduction to geophysical tomography, First Break, **2**, 20-26.